

## Genome analysis

# CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation

Honglei Liu<sup>1,2</sup>, Zheng Wei<sup>1</sup>, Antonia Dominguez<sup>3</sup>, Yanda Li<sup>1</sup>,  
Xiaowo Wang<sup>1,\*</sup> and Lei S. Qi<sup>2,3,4,\*</sup>

<sup>1</sup>Bioinformatics Division, Center for Synthetic and Systems Biology, TNLIST/Department of Automation, Tsinghua University, Beijing 100084, China, <sup>2</sup>Stanford Chemistry, Engineering & Medicine for Human Health (ChEM-H), <sup>3</sup>Department of Bioengineering and <sup>4</sup>Department of Chemical and Systems Biology, Stanford University, 443 Via Ortega, Shriram Center 376, Stanford, CA 94305-4125, USA

\*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on April 13, 2015; revised on July 3, 2015; accepted on July 16, 2015

## Abstract

**Summary:** The CRISPR/Cas9 system was recently developed as a powerful and flexible technology for targeted genome engineering, including genome editing (altering the genetic sequence) and gene regulation (without altering the genetic sequence). These applications require the design of single guide RNAs (sgRNAs) that are efficient and specific. However, this remains challenging, as it requires the consideration of many criteria. Several sgRNA design tools have been developed for gene editing, but currently there is no tool for the design of sgRNAs for gene regulation. With accumulating experimental data on the use of CRISPR/Cas9 for gene editing and regulation, we implement a comprehensive computational tool based on a set of sgRNA design rules summarized from these published reports. We report a genome-wide sgRNA design tool and provide an online web-site for predicting sgRNAs that are efficient and specific. We name the tool CRISPR-ERA, for clustered regularly interspaced short palindromic repeat-mediated eediting, repression, and activation (ERA).

**Availability and implementation:** <http://CRISPR-ERA.stanford.edu>.

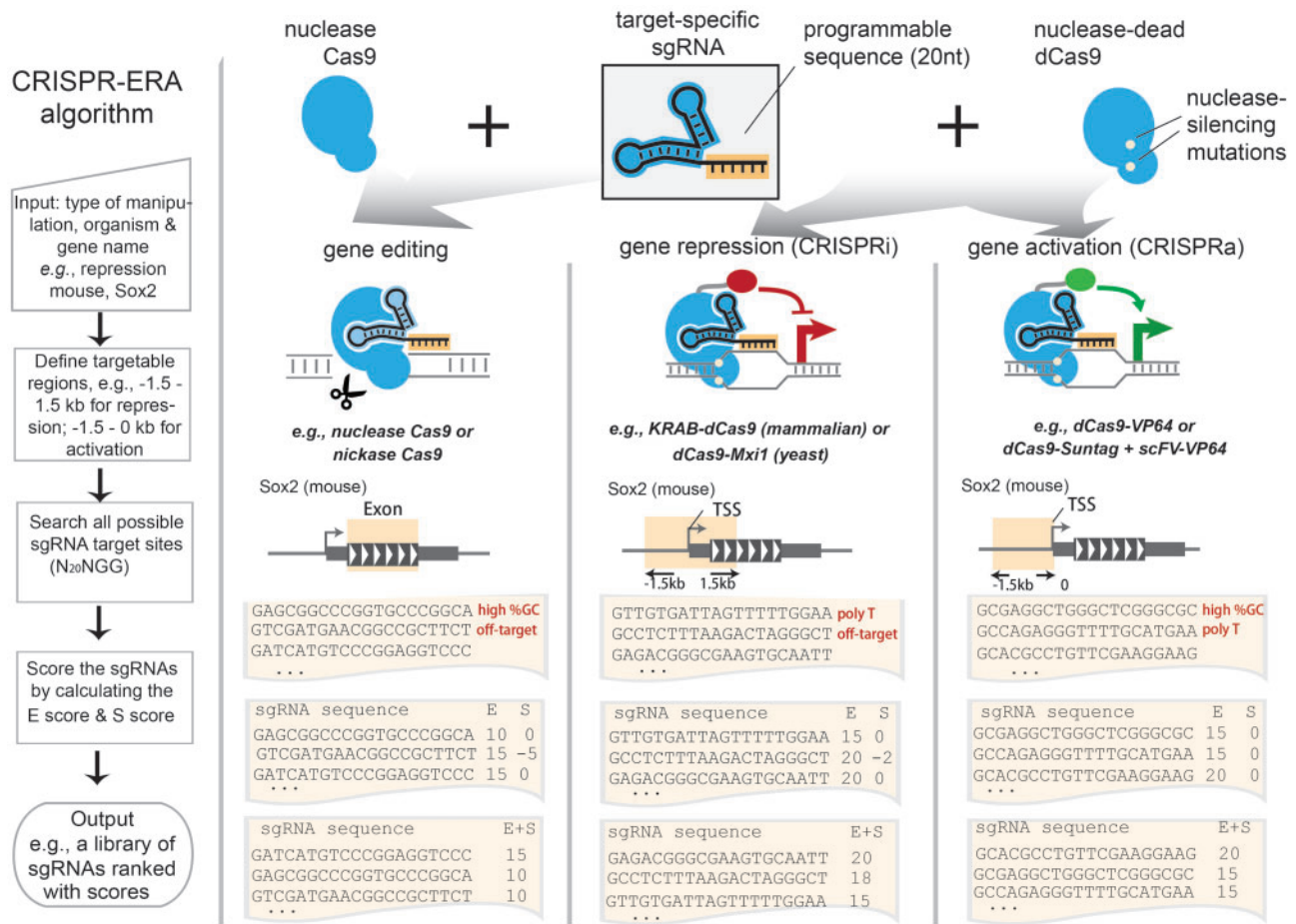
**Contact:** stanley.qi@stanford.edu or xwwang@tsinghua.edu.cn

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

## 1 Introduction

The bacterial adaptive immune system, CRISPR (clustered regularly interspaced short palindromic repeats), was recently developed as a powerful and multi-purpose technology for genome engineering, including editing (modifying the genomic sequence) (Cong *et al.*, 2013; Mali *et al.*, 2013), and regulation (repressing or activating expression of genes) (Gilbert *et al.*, 2013, 2014; Qi *et al.*, 2013). The system is highly programmable, utilizing a single protein, the nuclease Cas9 for editing or the nuclease-deficient dCas9 for regulation. A single guide RNA (sgRNA) is required for precise and programmable DNA targeting (Doudna and Charpentier, 2014). Effective

and specific genome engineering requires careful design of sgRNAs, which remains a major challenge. Computational tools have been used to facilitate the design of sgRNAs for CRISPR editing but not for other applications such as transcriptional regulation. These computational tools should enable automated sgRNA design and off-target site validation (Bae *et al.*, 2014; Doench *et al.*, 2014; Heigwer *et al.*, 2014; O'Brien and Bailey, 2014; Xiao *et al.*, 2014). A major goal of our designer tool is to address the discrepancy for designing sgRNAs that allow efficient and highly specific repression or activation of genes and for generating genome-wide sgRNA libraries for genetic screening in different organisms.



**Fig. 1.** CRISPR-ERA workflow and example. The CRISPR-ERA algorithm takes input information, including types of genome manipulation, organism, and gene name or genome location, and then computes and evaluates sgRNAs within the targeted genome region. By default, for editing, the algorithm chooses sgRNA sequences within coding region; for repression, the algorithm computes sgRNA binding sites within a 3 kb region centered at TSS (or the sense strand of the 5' end of the gene for bacteria only); for activation, the algorithm computes sgRNA binding sites up to 1.5 kb upstream of TSS. In this figure, mouse gene Sox2 is shown as an example. E, efficacy score; S, specificity penalty score (Supplementary Methods)

Here, we describe CRISPR-ERA webserver, an automated and comprehensive sgRNA design tool for CRISPR-mediated editing, repression, and activation (ERA) (Fig. 1). CRISPR-ERA utilizes a fast algorithm to search for genome-wide sgRNA binding sites and evaluates their efficiency and specificity using a set of rules summarized from published data for CRISPR editing, repression and activation (Cong *et al.*, 2013; Doudna and Charpentier, 2014; Gilbert *et al.*, 2014; Qi *et al.*, 2013; Ran *et al.*, 2013). The design features are annotated and the target sites can be visualized in a genome browser. We also provide a local version for the generation of whole-genome sgRNA libraries.

**2 Methods**

For each target gene or genomic site, CRISPR-ERA first searches all targetable sites in that particular organism for patterns of N<sub>20</sub>NGG (N = any nucleotide). Each target sequence is then calculated for two scores (Supplementary Methods): (i) an efficacy score (E-score) based on the sequence features such as GC content (%GC), presence of poly-thymidine (which is a terminator for effective transcription of sgRNAs), and location information such as the distance from target gene transcriptional start sites (TSS); and (ii) a specificity score (S-score) based on the genome-wide off-target

binding sites. For each sgRNA design, we compute the genome-wide sequences that contain an adjacent NRG (R = A or G) protospacer adjacent motif (PAM) site and zero, one, two, or three mismatches complementary to the sgRNA using Bowtie (Langmead *et al.*, 2009), which are regarded as off-target binding sites. The penalty score for NAG off-target is smaller than NGG off-target. The sgRNAs are finally ranked by the sum of E-score and S-score (Fig. 1; Supplementary Fig. S1).

We implement a user-friendly web server (<http://CRISPR-ERA.stanford.edu>) that hosts the web application for the sgRNA designer tool. The webserver will host a broad category of sequenced organisms. Currently, it provides sgRNA design service for nine most commonly used prokaryotic and eukaryotic organisms including *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Danio rerio*, *Rattus norvegicus*, *Mus musculus*, and *Homo sapiens*, etc. (Supplementary Table S1.). The web application enables rapid searching in the pre-assembled sgRNA database using an indexing searching approach (Supplementary Methods). The program outputs the sequence, target location, off-target details of the possible sgRNAs, their E- and S-scores etc. Results can be visualized using the UCSC genome browser to highlight the custom tracks (Supplementary Fig. S2 and S3).

### 3 Conclusions

CRISPR-ERA enables easy, fast, and predictive design of sgRNAs for broad applications of CRISPR in genome editing, gene repression and activation. The tool can be applied to other types of CRISPR applications such as genome imaging (Chen *et al.*, 2013) and CRISPR synthetic circuit design (Kiani *et al.*, 2014), and expanded to other organisms. We also provide the source code for the generation of whole-genome sgRNA libraries, useful for genome-wide screening based on CRISPR, CRISPRi or CRISPRa (Gilbert *et al.*, 2014, Shalem *et al.*, 2013; Wang *et al.*, 2013).

### Funding

L.S.Q. acknowledges support from NIH Office of The Director (OD), and National Institute of Dental & Craniofacial Research (NIDCR). H.L., Z.W., Y.L. and X.W. are supported by the 973 program (2012CB316503), NSFC (grant number 61322310, 31371341) and FANEDD (grant number 201158). This work was also supported by NIH Director's Early Independence Award (grant number OD017887, L.S.Q.) and NIH R01 (grant number DA036858, L.S.Q.).

*Conflict of Interest:* none declared.

### References

- Bae, S. *et al.* (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, **30**, 1473–1475.
- Chen, B. *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, **155**, 1479–1491.
- Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
- Doench, J.G. *et al.* (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.*, **32**, 1262–1267.
- Doudna, J.A. and Charpentier, E. (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096–1258096.
- Gilbert, L.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–451.
- Gilbert, L.A. *et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*, **159**, 647–661.
- Heigwer, F. *et al.* (2014) E-CRISP: fast CRISPR target site identification. *Nat. Methods*, **11**, 122–123.
- Kiani, S. *et al.* (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods*, **11**, 723–726.
- Langmead, B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
- Mali, P. *et al.* (2013) RNA-guided human genome Engineering via Cas9. *Science*, **339**, 823–826.
- O'Brien, A. and Bailey, T.L. (2014) GT-Scan: identifying unique genomic targets. *Bioinformatics*, **30**, 2673–2675.
- Qi, L.S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
- Ran, F.A. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **155**, 479–480.
- Shalem, O. *et al.* (2013) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
- Wang, T. *et al.* (2013) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
- Xiao, A. *et al.* (2014) CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, **30**, 1180–1182.